Redes Complexas

Uma introdução

Leandro Chaves Rêgo, Jesus Ossian Cunha, Pablo Ignacio Fierens

Universidade Federal do Ceará

05 e 06 de Novembro de 2024

Conteúdo

Métricas de Centralidade

Conceito de centralidade

Quem são os principais vértices da rede?

- Qual o aeroporto mais importante em uma rede representando a malha aérea de aviões?
- Como identificar os perfis mais importantes em uma rede social para apresentar um determinado anúncio?

Métricas de centralidade

Fazem parte de um grupo de métricas de redes chamadas métricas locais, que descrevem características individuais dos nós.

Métricas Geométricas

Conceito

- Associa a centralidade ou importância de um nó com respeito a distância geodésica deste nó em relação aos demais nós da rede.
- Em grafos direcionados, existem duas opções: diferem se consideramos as distâncias a partir do nó ou distâncias para chegar no nó que estamos calculando a centralidade.
- Para simplificar, iremos descrever de acordo com a primeira opção.

Métricas geométricas

Excentricidade

Seja $G_i = (V_i, E_i)$ a componente da rede que contém o nó i. A excentricidade do nó i é dada por:

$$ec(i) = \max_{j \in V_i} d(i, j).$$

Para garantir que os nós mais centrais estejam mais próximos dos outros, a centralidade de excentricidade do nó i é dada pelo recíproco da sua excentricidade:

$$C_{ec}(i) = \frac{1}{ec(i)}.$$

Excentricidade

Exemplo

Em um grafo completo todos os nós possuem excentricidade e centralidade de excentricidade iguais a 1. Em um círculo com $2n$ vértices, todos os nós tem excentricidade igual a n e centralidade de excentricidade igual a $1/n$. Em uma linha com n vértices, a excentricidade das folhas é $n-1$ e a centralidade de excentricidade é $1/(n-1)$.

Métricas geométricas

Proximidade

É igual ao recíproco da média aritmética das distâncias geodésicas do nó para os demais nós da mesma componente do nó, ou seja:

$$C_c(i) = \frac{||V_i|| - 1}{\sum_{j \in V_i - \{i\}} d(i,j)}.$$

Centralidade de Proximidade

Exemplo

Em um círculo com $2n$ nós, a centralidade de proximidade de qualquer nó i é dada por:

$$\frac{2n-1}{n+\sum_{j=1}^{n-1} 2j} = \frac{2n-1}{n^2}.$$

Por outro lado, em um círculo com $2n+1$ nós, a centralidade de proximidade de qualquer nó i é dada por:

$$\frac{2n}{\sum_{j=1}^{n} 2j} = \frac{2n}{n(n+1)} = \frac{2}{n+1}.$$

Métricas geométricas

Harmônica

Recíproco da média harmônica das distâncias:

$$C_h(i) = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d(i,j)},$$

em que n é o número total de nós na rede.
Bem definida em redes com mais de uma componente em que existem nós que não se conectam, e portanto, têm distâncias geodésicas infinitas.

Centralidade Harmônica

Exemplo

Vamos considerar uma rede que é formada por duas componentes sendo uma correspondente a um grafo estrela com n_1 nós e outra correspondente a um círculo com n_2 nós. Então, a centralidade harmônica de um nó periférico da estrela é:

$$\frac{1}{n_1+n_2-1}\left(1+(n_1-2)\frac{1}{2}\right)=\frac{n_1}{2(n_1+n_2-1)}.$$

A centralidade harmônica do nó central da estrela é:

$$\frac{1}{n_1+n_2-1}(n_1-1).$$

Finalmente, a centralidade harmônica dos nós no círculo serão:

$$\begin{cases} \frac{1}{n_1+n_2-1}\left(\frac{2}{n_2}+\sum_{k=1}^{\frac{n_2}{2}-1}\frac{2}{k}\right) & \text{se } n_2 \text{ for par;} \\ \frac{1}{n_1+n_2-1}\left(\sum_{k=1}^{\frac{n_2-1}{2}}\frac{2}{k}\right) & \text{se } n_2 \text{ for ímpar.} \end{cases}$$

Métricas geométricas

Média-p

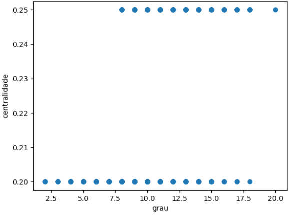
$$C_p(i) = \begin{cases} \left(\frac{\sum_{j\in V_i-\{i\}} d(i,j)^p}{n-1}\right)^{-\frac{1}{p}}, & \text{se } p \neq 0, \\ \left(\prod_{j\in V_i-\{i\}} d(i,j)\right)^{-\frac{1}{n-1}}, & \text{se } p = 0. \end{cases}$$

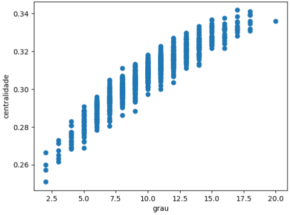
- $C_0(i) = \lim_{p\to 0} C_p(i)$: média geométrica.
- $C_{-1}(i) = C_h(i)$.
- $C_1(i) = C_c(i) \times \frac{n-1}{||V_i||-1}$.
- $\lim_{p\to +\infty} C_p(i) = C_{ec}(i)$.

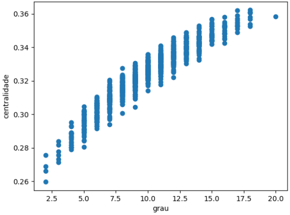
Métricas geométricas: Exemplo I

```python
import matplotlib.pyplot as plt
import numpy as np
import networkx as nx

plt.close('all')
#grafo aleatório de Erdös—Rényi com 1000 nós e probabilidade 0.01
ger = nx.gnp_random_graph(1000,0.01)
graus = dict(nx.degree(ger)).values()

#centralidade de excentricidade
ec  = nx.eccentricity(ger)
cec = [1.0/ec[n] for n in ec]

#centralidade de proximidade
#só salvamos os valores do dicionário e transformamos em lista
cc  = list(nx.closeness_centrality(ger).values())

#centralidade harmônica
#só salvamos os valores do dicionário e transformamos em lista
chu= list(nx.harmonic_centrality(ger).values())
ch = [x/(len(chu)-1) for x in chu]  #normalização
```

Métricas geométricas: Exemplo II

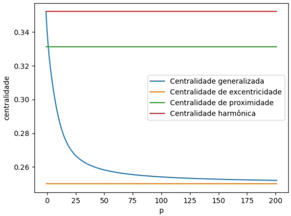


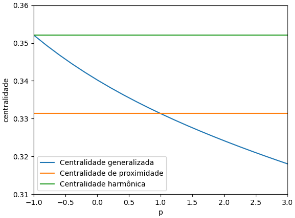




Média-p: Exemplo I

```
#valores de p
p = np.linspace(-1,201,1000)
#sp = comprimento dos caminhos mais curtos desde nó 1
sp= np.array(list(nx.shortest_path_length(ger,1).values()))
#cp = centralidade generalizada para o nó 1
cp= np.zeros_like(p)
for k in range(p.shape[0]):
    if p[k]:
        cp[k] = np.mean(sp[1:]**p[k])**(-1.0/p[k])
    else:
        cp[k] = np.prod(sp[1:])**(-1.0/(sp.shape[0]-1))
```





Centralidade de Intermediação

Ideia

Quantifica o quanto os vértices são capazes de atuar como intermediários entre outros dois vértices, podendo portanto controlar o fluxo de informação entre eles.

Definição

$$C_b(i) = \frac{1}{(n-1)(n-2)} \sum_{(j,k):j\neq k, i\notin\{j,k\}} \frac{Q_{j,k}(i)}{Q_{j,k}},$$

onde $Q_{j,k}$ é o número de caminhos geodésicos iniciando no vértice j e terminando no vértice k e que $Q_{j,k}(i)$ é o número de caminhos geodésicos que iniciam em j, terminam em k e passam pelo vértice i.

Centralidade de Intermediação

Exemplo

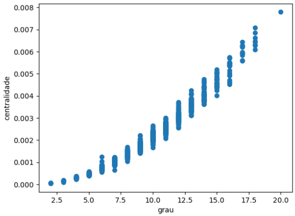
Considere novamente a rede que consiste de uma estrela de n_1 nós e um círculo de n_2 nós, mas na qual adicionamos uma ligação entre o nó central da estrela e um nó do círculo. Note que qualquer caminho geodésico partindo de (resp., terminando em) um nó periférico da estrela passa pelo vértice central da estrela. Deste modo, a centralidade de intermediação do nó central da estrela é:

$$
\begin{aligned}
& \frac{1}{(n_1+n_2-1)(n_1+n_2-2)}[(n_1-1)(n_1+n_2-2)+n_2(n_1-1)] \\
= \quad & \frac{(n_1-1)(n_1+2n_2-2)}{(n_1+n_2-1)(n_1+n_2-2)}.
\end{aligned}
$$

Além disso, os nós periféricos da estrela possuem centralidade de intermediação nula.

Centralidade de Intermediação: Exemplo I

```
#centralidade de intermediação
#só salvamos os valores do dicionário e transformamos em lista
cb = list(nx.betweenness_centrality(ger,normalized=True).values())
plt.figure()
plt.plot(graus,cb,'o')
plt.xlabel('grau')
plt.ylabel('centralidade')
```



Centralidade de Autovetor

Definição

$$C_{eg}(i) = \frac{1}{\lambda_1} \sum_{j=1}^{n} A(i,j) C_{eg}(j),$$

em que λ_1 o maior autovalor de A.
Em redes direcionadas, temos duas opções para a centralidade de autovetor, a depender se um vértice recebe importância dos vértices que estão ligados a ele ou dos vértices aos quais ele está ligado. Em geral, temos a primeira situação. É comum normalizar o vetor de centralidades de autovetor de modo que este tenha normal euclideana 1.

Problemas

Em redes direcionadas, nós que possuem grau de entrada nulo possuem centralidade de autovetor nula.
Todos os nós em uma rede direcionada acíclica possuem centralidade de autovetor nula.

Centralidade de Katz

Ideia

A centralidade de Katz resolve o problema adicionando um valor pequeno de centralidade para todos os nós da rede:

$$C_K(i) = \alpha \sum_{j \neq i} A(i,j) C_K(j) + \beta,$$

em que α e β são constantes positivas. β é o valor de centralidade que os nós de grau nulo possuem e $\alpha < 1/\lambda_1$ em que λ_1 é o maior autovalor de A. Na prática, é comum adotar-se o valor de α próximo a esse valor máximo e com um pequeno valor de β.

PageRank

Ideia

De acordo com a centralidade de Katz, o fato de que algum vértice importante está ligado a um dado vértice dá alta importância a este dado vértice, independente de quantas ligações na rede possui aquele vértice importante.
O PageRank assume que a importância de um nó é divida igualmente entre os seus vizinhos para os quais ele possui ligação. Deste modo:

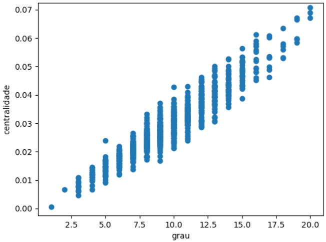
$$C_{PR}(i) = (1-\alpha) + \alpha \sum_{j \neq i} \frac{A(j,i) C_{PR}(j)}{d^{out}(j)},$$

em que $\alpha \in (0,1)$ (usualmente, $\alpha = 0{,}85$). O PageRank é utilizado pelo Google. α pode ser pensado como a probabilidade de um usuário continuar navegando entre páginas através dos links. No caso de um vértice j com $d^{out}(j)$ nulo, como j não contribui para a importância de nenhum outro nó, adotaremos a convenção de que $d^{out}(j) = 1$, de modo que o termo $A(j,i)/d^{out}(j)$ seja nulo.

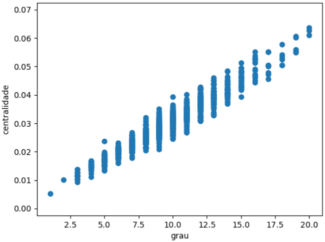
Autovetor & Katz: Exemplo I

```
import numpy as np
import networkx as nx

#grafo aleatório de Erdös-Rényi com 1000 nós e probabilidade
    0.01
ger = nx.gnp_random_graph(1000,0.01)
graus = dict(nx.degree(ger)).values()

#centralidade de autovetor
cev = list(nx.eigenvector_centrality_numpy(ger).values())
#centralidade de Katz
ck1 = list(nx.katz_centrality_numpy(ger,alpha=0.08,beta=1).
    values())
ck2 = list(nx.katz_centrality_numpy(ger,alpha=0.01,beta=1).
    values())
ck3 = list(nx.katz_centrality_numpy(ger,alpha=0.08,beta=2).
    values())
```

Autovetor & Katz: Exemplo II

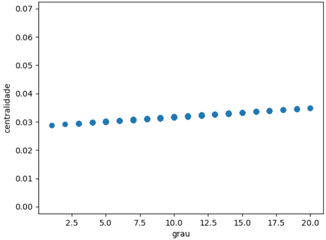


(a) Centralidade de autovetor.

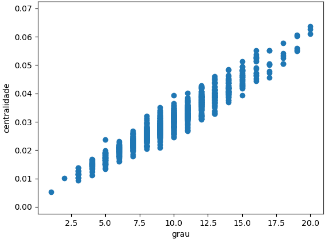


(b) Katz com $\alpha = 0.08$ e $\beta = 1$.

Autovetor & Katz: Exemplo III



(a) Centralidade de Katz com $\alpha = 0.01$ e $\beta = 1$.

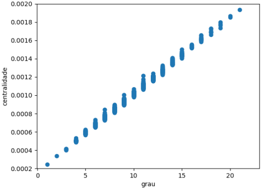


(b) Centralidade de Katz com $\alpha = 0.08$ e $\beta = 2$.

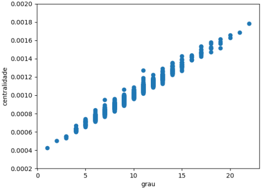
PageRank: Exemplo I

```python
import networkx as nx

#grafo aleatório de Erdös-Rényi com 1000 nós e probabilidade
     0.01
ger = nx.gnp_random_graph(1000,0.01)
graus = dict(nx.degree(ger)).values()

#centralidade de PageRank
cpr1 = list(nx.pagerank(ger,alpha=0.85).values())
cpr2 = list(nx.pagerank(ger,alpha=0.65).values())
cpr3 = list(nx.pagerank(ger,alpha=0.25).values())
```

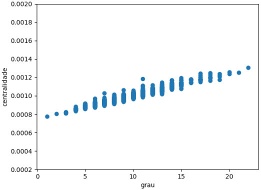
PageRank: Exemplo II



(a) $\alpha = 0.85$.



(b) $\alpha = 0.65$.



(c) $\alpha = 0.25$.

Coeficiente de Agrupamento Local

Definição

Permite avaliar o quanto os nós são capazes de proporcionar interação entre os seus vizinhos. O coeficiente de agrupamento local de um nó i é dado por:

$$cl(i) = \frac{\sum_{(j,k):j \neq i, k \neq i, k \neq j} A(i,j)A(i,k)A(j,k)}{\sum_{(j,k):j \neq i, k \neq i, k \neq j} A(i,j)A(i,k)}.$$

Pode-se definir o coeficiente de agrupamento local médio:

$$\overline{cl}(G) = \frac{1}{n}\sum_{i=1}^{n} cl(i).$$

Coeficiente de Agrupamento Local: Exemplo I

Considere uma rede que consiste de um círculo com $n-1$ nós todos conectados a um nó central. Pode-se mostrar que:

$$\overline{cl}(G) = \frac{1}{n}\left[\frac{2(n-1)}{3} + \frac{2}{n-2}\right] = \frac{2n^2-6n+10}{3n(n-2)}.$$

e

$$cl(G) = \begin{cases} 1 & \text{se } n = 4, \\ \frac{2(n-1)+(n-1)}{3(n-1)+\frac{(n-1)(n-2)}{2}} = \frac{6}{n+4} & \text{se } n > 4. \end{cases}$$

Coeficiente de Agrupamento Local: Exemplo II

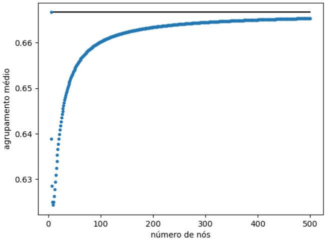
Em uma rede em que todos os vértices têm grau no mínimo igual a 2, pode-se mostrar que:

$$cl(G) = \sum_{i=1}^{n} cl(i) \frac{d(i)(d(i)-1)/2}{\sum_{j=1}^{n} d(j)(d(j)-1)/2},$$

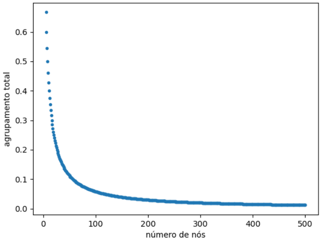
em que $cl(G)$ é o coeficiente de agrupamento total e $cl(i)$ é o coeficiente de agrupamento local do vértice i.

```python
import networkx as nx
import numpy as np

ccla = []
cclt = []
n = np.linspace(5,500,496)
for k in n:
    gci = nx.wheel_graph(int(k))
    ccla.append(nx.average_clustering(gci))
    cclt.append(nx.transitivity(gci))
```

Coeficiente de Agrupamento Local: Exemplo III



(a) Agrupamento médio.



(b) Agrupamento total.

Homofilia

Ideia

Uma das informações importantes em redes complexas e, em especial, em redes sociais é saber se os nós possuem alguma tendência se ligar mais a nós similares a si mesmo, o que é conhecido como homofilia ou associatividade. Por exemplo, cientistas tendem a ter contato com mais cientistas que, em geral, outras pessoas possuem contatos com cientistas. Pessoas tendem a ter mais relacionamentos intrarraciais que inter-raciais.

Índice EI

Definição

Considere um grafo particionado em grupos $\{V^1, V^2, \ldots, V^a\}$ de acordo com atributos que esses nós possuam. O índice EI mede a homofilia através de uma razão que envolve a quantidade de ligações entre nós de grupos distintos (EL) e a quantidade de ligações entre nós do mesmo grupo (IL). Ele pode ser calculado para um subconjunto qualquer de nós da rede V':

$$EI(V') = \frac{EL(V') - IL(V')}{EL(V') + IL(V')}.$$

O valor do índice EI varia no intervalo $[-1, 1]$, em que um índice EI igual a -1 (resp., 1) indica um subconjunto de nós que só possui ligações internas (resp., externas).

Índice EI

Problema

- O valor do índice EI é afetado pelo fato de um certo grupo de atributos ser mais ou menos frequente na rede que outros.
    - Suponha que um indivíduo faça parte de algum grupo minoritário.
    - Pode acontecer que este indivíduo possua mais ligações externas do que internas no grupo e, portanto, tenha um índice EI positivo.
    - Isso não necessariamente quer dizer que este indivíduo tenha preferência por ter ligações externas.
- Precisaríamos comparar com o número esperado de ligações que seria obtido em uma rede aleatória com o mesmo conjunto de nós e o mesmo número de arestas escolhidas aleatoriamente.

Assortatividade

Definição

A fração de arestas que ligam um nó no conjunto V^i a um nó no conjunto V^j:

$$e_{ij} = \frac{1}{c} \sum_{k_1 \in V^i} \sum_{k_2 \in V^j} A(k_1, k_2),$$

em que $c = m$ ou $c = 2m$ se a rede é direcionada ou não-direcionada, respectivamente.
A fração de arestas que iniciam (resp., terminam) em vértices do conjunto V_i é dada, respetivamente, por $a_i = \sum_j e_{ij}$, $b_i = \sum_j e_{ji}$.
O coeficiente de assortatividade é dado por:

$$As^c(G) = \frac{\sum_{i=1}^{a} e_{ii} - \sum_{i=1}^{a} a_i b_i}{1 - \sum_{i=1}^{a} a_i b_i},$$

em que a normalização no denominador é para garantir que, no caso de homofilia perfeita, o coeficiente tenha valor igual a 1.

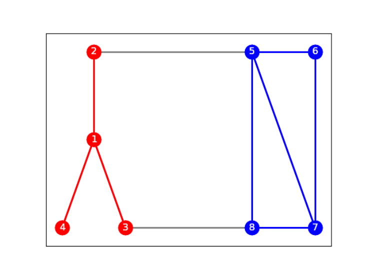
Homofilia

Exemplo

Considere a rede da Figura abaixo. Para calcular o valor do índice EI da rede, temos: $EI = \frac{2-8}{2+8} = -0.6$.
Já o valor da assortatividade categórica é dado por:

$$As^c(G) = \frac{6/20 + 10/20 - (8/20)^2 - (12/20)^2}{1 - (8/20)^2 - (12/20)^2} = 0.583.$$

Ambas as métricas indicam uma rede com homofilia.



Assortatividade

Definição para atributos escalares

O coeficiente de assortatividade para atributos escalares é calculado como o coeficiente de correlação de Pearson. Seja x_i o atributo escalar do no i. Como o nó i possui $d(i)$ arestas ligadas a ele, o valor médio dos atributos escalares ao longo das arestas é dado por:

$$\overline{x} = \frac{\sum_i d(i) x_i}{\sum_i d(i)} = \frac{1}{2m} \sum_i d(i) x_i.$$

O coeficiente de assortatividade é:

$$As^e(G) = \frac{\frac{1}{2m} \sum_{ij} \left(A(i,j) x_i x_j \right) - \overline{x}^2}{\frac{1}{2m} \sum_{ij} \left(A(i,j) x_i^2 \right) - \overline{x}^2}.$$

Assortatividade de grau

Caso especial

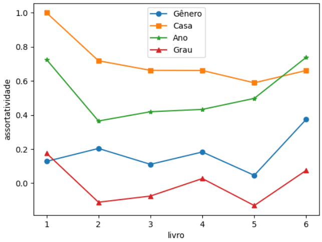
- Um caso especial de assortatividade com atributo escalar que é bastante utilizado é quando o atributo escalar é a própria centralidade de grau, ou seja, $x_i = d(i)$.
- Esta assortatividade mede se na rede existe uma tendência de nós muito conectados estarem conectados entre si e nós pouco conectados estarem ligados entre si.
- Redes com assortatividade de grau positiva possuem nós bastante conectados ligados entre si, formando um núcleo da rede que se destaca da periferia. Essa tendência é bastante observada em redes sociais.

Assortatividade: Exemplo I

```python
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt

def carregarlivro(livro):
    atri = pd.read_csv('hpattributes.txt', sep='\t') #sem encabeçamento
    ares = pd.read_csv('hpbook{:1d}.txt'.format(livro), sep=' ', header=None)
    nome = pd.read_csv('hpnames.txt', sep='\t')

    gpotter = nx.DiGraph()
    n = atri.shape[0]
    #Primeiro agregamos os nós
    for k in range(n):
        gpotter.add_node(k,nome=nome['name'][k],
                         ano=atri['schoolyear'][k],
                         gen=atri['gender'][k],
                         casa=atri['house'][k])
    #Construímos a rede
    for k in range(n):
        for m in range(n):
            if ares.values[k][m] == 1:
                gpotter.add_edge(k,m)

    agen = nx.attribute_assortativity_coefficient(gpotter, "gen")
    acasa= nx.attribute_assortativity_coefficient(gpotter, "casa")
    aano = nx.numeric_assortativity_coefficient(gpotter.to_undirected(), "ano")
    agrau= nx.degree_assortativity_coefficient(gpotter.to_undirected())

    return agen,acasa,aano,agrau
```

Assortatividade: Exemplo II

```
agen =  6*[0]
acasa=  6*[0]
aano =  6*[0]
agrau=  6*[0]
livros = [1,2,3,4,5,6]
for k in range(len(livros)):
    agen[k],acasa[k],aano[k],agrau[k]=carregarlivro(livros[k])
```



Obrigado!

Contatos:

Leandro Rêgo - leandro@dema.ufc.br
Jesus Ossian - jesus.ossian@dema.ufc.br
Pablo Fierens - pfierens@itba.edu.ar